CS230/561: Probability and Statistics for Computer Science

Clint P. George

August 28, 2024

Some slides/examples are motivated from *MIT 6-012 Introduction to Probability* by J. Tsitsiklis

Topic outline

1. Consequences of axioms
2. Probability calculation: examples
3. Conditional probability
4. The Bayes theorem
5. Independence of events

Probability axioms

Let Ω be the sample space and $A, B \subset \Omega$

1. Non negativity, $P(A) \geq 0$
2. Normalization, $P(\Omega) = 1$
3. Finite additivity: if A, B are disjoint (mutually exclusive), then

$$P(A \cup B) = P(A) + P(B)$$

Consequences of axioms

Probability axioms: consequences

A $P(\phi) = 0$

B $P(A) + P(A^{\complement}) = 1$

C $P(A) \leq 1$, for any $A \subset \Omega$

D $P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$, assuming $A_i \subset \Omega$ are disjoint

Finite additivity: a use case

- Probability of a finite set $\{s_1, s_2, \ldots, s_k\}$—a set which we can count and finish counting

$$P(\{s_1, s_2, \ldots, s_k\}) = P(\{s_1\} \cup \{s_2\} \cup \ldots \cup \{s_k\})$$

Finite additivity: a use case

- Probability of a finite set $\{s_1, s_2, \ldots, s_k\}$—a set which we can count and finish counting

$$
\begin{aligned}
P(\{s_1, s_2, \ldots, s_k\}) &= P(\{s_1\} \cup \{s_2\} \cup \ldots \cup \{s_k\}) \\
&= P(\{s_1\}) + P(\{s_2\}) + \ldots + P(\{s_k\})
\end{aligned}
$$

Finite additivity: a use case

- Probability of a finite set $\{s_1, s_2, \ldots, s_k\}$—a set which we can count and finish counting

$$
\begin{aligned}
P(\{s_1, s_2, \ldots, s_k\}) &= P(\{s_1\} \cup \{s_2\} \cup \ldots \cup \{s_k\}) \\
&= P(\{s_1\}) + P(\{s_2\}) + \ldots + P(\{s_k\}) \\
&= P(s_1) + P(s_2) + \ldots P(s_k)
\end{aligned}
$$

Example: rolling a dice

- Sample space: $\{1, 2, 3, 4, 5, 6\}$

Example: rolling a dice

- Sample space: $\{1, 2, 3, 4, 5, 6\}$
- We assume we have an unbiased dice, i.e., all sides are equally likely

$$P(\{1\}) = P(\{2\}) = \ldots = \frac{1}{6}$$

Example: rolling a dice

- Sample space: $\{1, 2, 3, 4, 5, 6\}$
- We assume we have an unbiased dice, i.e., all sides are equally likely

$$P(\{1\}) = P(\{2\}) = \ldots = \frac{1}{6}$$

We can generalize this to a sample space of n elements.

Example: rolling a dice

- Sample space: $\{1, 2, 3, 4, 5, 6\}$
- We assume we have an unbiased dice, i.e., all sides are equally likely

$$P(\{1\}) = P(\{2\}) = \ldots = \frac{1}{6}$$

  We can generalize this to a sample space of n elements.

- What is the probability of rolling an even number

Example: rolling a dice

- Sample space: $\{1, 2, 3, 4, 5, 6\}$
- We assume we have an unbiased dice, i.e., all sides are equally likely

$$P(\{1\}) = P(\{2\}) = \ldots = \frac{1}{6}$$

We can generalize this to a sample space of n elements.

- What is the probability of rolling an even number

$$P(\{2, 4, 6\}) = P(\{2\}) + P(\{4\}) + P(\{6\})$$

based on *finite additivity* axiom.

Union bound

- Let A, B are not disjoint events, i.e., $A \cap B \neq \phi$, then we have

$$P(A \cup B) \leq P(A) + P(B)$$

Finite additivity: another use case

$P(A \cup B \cup C) =$

Finite additivity: another use case

$$P(A \cup B \cup C) = P(A) + P(A^{\complement} \cap B) + P(A^{\complement} \cap B^{\complement} \cap C)$$

Identify partitions—i.e., disjoint sets; we can then apply the additivity rule.

Probability calculation: examples

Discrete example: rolling a dice

- *Two rolls* of a 4-sided dice; outcomes represented by variables X, Y
- All outcomes are equally likely and have probability

$$\frac{1}{16}$$

Discrete example: rolling a dice

- *Two rolls* of a 4-sided dice; outcomes represented by variables X, Y
- All outcomes are equally likely and have probability

$$\frac{1}{16}$$

- What is the probability $P(X = 1)$?

Discrete example: rolling a dice

- *Two rolls* of a 4-sided dice; outcomes represented by variables X, Y
- All outcomes are equally likely and have probability

$$\frac{1}{16}$$

- What is the probability $P(X = 1)$?
- What is the probability $P(Z = 4), P(Z = 2)$, where $Z = \min(X, Y)$?

Discrete uniform law

- Assume finite set Ω consists of n equally likely elements, e.g., outcomes of dice rolling
- Probability of observing an element in the set is given by $1/n$

Discrete uniform law

- Assume finite set Ω consists of n equally likely elements, e.g., outcomes of dice rolling
- Probability of observing an element in the set is given by $1/n$
- Assume $A \subset \Omega$ consists of k elements
- What is the probability $P(A)$?

Discrete uniform law

- Assume finite set Ω consists of n equally likely elements, e.g., outcomes of dice rolling
- Probability of observing an element in the set is given by $1/n$
- Assume $A \subset \Omega$ consists of k elements
- What is the probability $P(A)$?

The probability of a finite set is the sum of the probabilities of finite elements of the set.

Continuous example: unit square

- Let (x, y) be the sample such that $0 \leq x, y \leq 1$
- Probability law: area of the set

Continuous example: unit square

- Let (x, y) be the sample such that $0 \leq x, y \leq 1$
- Probability law: area of the set

- Let $A = \{(x, y) :\ x + y \leq 1/2\}$ be subset

Continuous example: unit square

- Let (x, y) be the sample such that $0 \leq x, y \leq 1$
- Probability law: area of the set

- Let $A = \{(x, y) : x + y \leq 1/2\}$ be subset

- What is $P(A)$?

Continuous example: unit square

- Let (x, y) be the sample such that $0 \leq x, y \leq 1$
- Probability law: area of the set
- Let $A = \{(x, y) : x + y \leq 1/2\}$ be subset
- What is $P(A)$?
- What is $P(\{3.1, 2.0\})$?

Probability computation: steps

- Define the sample space
- Define the probability law, e.g., the area of the set
- Identify an event of interest
- Compute

Additional axiom: countable additivity

If $A_1, A_2, \ldots$ is an infinite *sequence* of disjoint events,

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$$

[1] Georg Cantor (1845-1918) proved that the set of real numbers is uncountably infinite.

Additional axiom: countable additivity

If $A_1, A_2, \ldots$ is an infinite *sequence* of disjoint events,

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$$

- Additivity holds for *countable sequence of events*
- Unit square, real line[1], etc., are not countable—i.e., the elements cannot be arranged in a sequence

[1] Georg Cantor (1845-1918) proved that the set of real numbers is uncountably infinite.

Probability calculation: discrete but infinite Ω

- Experiment: toss a coin until you get the head
- Sample space

$$\Omega = \{1, 2, 3, \ldots\}$$

Probability calculation: discrete but infinite Ω

- Experiment: toss a coin until you get the head
- Sample space

$$\Omega = \{1, 2, 3, \ldots\}$$

- The probability of every outcome is given by

$$P(n) = \frac{1}{2^n}; \quad n = 1, 2, \ldots$$

- Check whether $P(\Omega) = 1$

Probability calculation: discrete but infinite Ω

- Experiment: toss a coin until you get the head
- Sample space

$$\Omega = \{1, 2, 3, \ldots\}$$

- The probability of every outcome is given by

$$P(n) = \frac{1}{2^n}; \quad n = 1, 2, \ldots$$

- Check whether $P(\Omega) = 1$
- Event: the outcome is even, i.e., $E = \{2, 4, \ldots\}$, what is $P(E)$?

Probability calculation: discrete but infinite Ω

- Experiment: toss a coin until you get the head
- Sample space

$$\Omega = \{1, 2, 3, \ldots\}$$

- The probability of every outcome is given by

$$P(n) = \frac{1}{2^n}; \quad n = 1, 2, \ldots$$

- Check whether $P(\Omega) = 1$
- Event: the outcome is even, i.e., $E = \{2, 4, \ldots\}$, what is $P(E)$?—apply *countable additivity*

Conditional probability

Conditioning — intuition

- Consider we have equally likely outcomes in the sample space Ω
- Events $A, B \subset \Omega$, $A \cap B \neq \phi$

Conditioning — definition

- Let $P(B) > 0$
- We define the conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Example: two rolls of 4-sided dice

- Let B be the event $\min(X, Y)$
- Let A be the event $\max(X, Y)$
- What is $P(A = 1|B = 2)$?
- What is $P(A = 3|B = 2)$?

Properties of conditional probabilities

Given non-empty event B, we have

- Nonnegativity, $P(A|B) \geq 0$

Properties of conditional probabilities

Given non-empty event B, we have

- Nonnegativity, $P(A|B) \geq 0$
- Nomalization, $P(\Omega|B) = 1$.

Properties of conditional probabilities

Given non-empty event B, we have

- Nonnegativity, $P(A|B) \geq 0$
- Nomalization, $P(\Omega|B) = 1$.

Proof. apply the definition of conditional probability.

Properties of conditional probabilities

Given non-empty event B, we have

- Nonnegativity, $P(A|B) \geq 0$
- Nomalization, $P(\Omega|B) = 1$.

  Proof. apply the definition of conditional probability.
- Finitie additivity, If $A \cap C = \phi$, then is $P(A \cup C|B) = P(A|B) + P(C|B)$

Properties of conditional probabilities

Given non-empty event B, we have

- Nonnegativity, $P(A|B) \geq 0$
- Nomalization, $P(\Omega|B) = 1$.

  Proof. apply the definition of conditional probability.

- Finitie additivity, If $A \cap C = \phi$, then is $P(A \cup C|B) = P(A|B) + P(C|B)$

  We can generalize this to finitely many or countable events

Conditional probability calculation: example

- Discrete events
- Event A: *airplane flying*
- Event R: *radar beacon*

Conditional probability calculation: example

- Discrete events
- Event A: *airplane flying*
- Event R: *radar beacon*

- What is $P(A \cap R)$?

Conditional probability calculation: example

- Discrete events
- Event A: *airplane flying*
- Event R: *radar beacon*

- What is $P(A \cap R)$?
- What is $P(R)$?

Conditional probability calculation: example

- Discrete events
- Event A: *airplane flying*
- Event R: *radar beacon*

- What is $P(A \cap R)$?
- What is $P(R)$?
- What is $P(A|R)$?

Multiplication rule

Conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Multiplication rule

Conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

2 events

$$P(A \cap B) = P(B)P(A|B)$$

Multiplication rule

Conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

2 events

$$\begin{aligned} P(A \cap B) &= P(B)P(A|B) \\ &= P(A)P(B|A) \end{aligned}$$

Multiplication rule

Conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

2 events

$$\begin{aligned} P(A \cap B) &= P(B)P(A|B) \\ &= P(A)P(B|A) \end{aligned}$$

3 events

$$\begin{aligned} P(A \cap B \cap C) &= P(A \cap B)P(C|A \cap B) \\ &= P(A)P(B|A)P(C|A \cap B) \end{aligned}$$

Multiplication rule

Conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

2 events

$$\begin{aligned} P(A \cap B) &= P(B)P(A|B) \\ &= P(A)P(B|A) \end{aligned}$$

3 events

$$\begin{aligned} P(A \cap B \cap C) &= P(A \cap B)P(C|A \cap B) \\ &= P(A)P(B|A)P(C|A \cap B) \end{aligned}$$

Multiplication rule

We can generalize the rule for k events $A_1, A_2, \ldots, A_k$

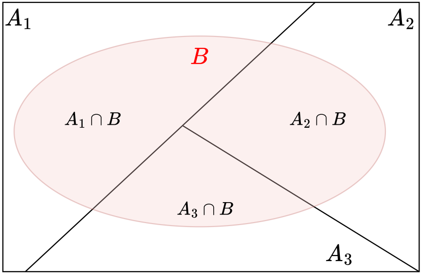
$$P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1) \prod_{i=2}^{k} P(A_i | \underbrace{A_1 \cap A_2 \cap \ldots \cap A_{i-1}}_{\text{all the previous events}})$$

Total probability theorem

- Total probability of an event $B \subset \Omega$ that occur at different scenarios
- Partition the sample space into disjoint events $A_1, A_2, \ldots, A_k \subset \Omega$

Total probability theorem

- Total probability of an event $B \subset \Omega$ that occur at different scenarios
- Partition the sample space into disjoint events $A_1, A_2, \ldots, A_k \subset \Omega$



(Example with three disjoint sets)

Total probability theorem

- Total probability of an event $B \subset \Omega$ that occur at different scenarios
- Partition the sample space into disjoint events $A_1, A_2, \dots, A_k \subset \Omega$

- For every event/scenario i, we have
  - $P(A_i)$ — probability of the scenario
  - $P(B|A_i)$ — probability of the event under that scenario

Total probability theorem

- Total probability of an event $B \subset \Omega$ that occur at different scenarios
- Partition the sample space into disjoint events $A_1, A_2, \ldots, A_k \subset \Omega$

- For every event/scenario i, we have
  - $P(A_i)$ — probability of the scenario
  - $P(B|A_i)$ — probability of the event under that scenario

- We then have

$$P(B) = \sum_{i=1}^{k} P(A_i) P(B|A_i)$$

Total probability theorem

- The total probability of an outcome that can be achieved through multiple events is the sum of the probabilities of those other events in the sample space.
- This law relates marginal probabilities to conditional probabilities.

- One can see a marginal probability as a weighted average of conditional probabilities, where the weights are determined by the likelihood of each case.

The Bayes theorem

The Bayes theorem

- Partition the sample space into disjoint events $A_1, A_2, \ldots, A_k \subset \Omega$

The Bayes theorem

- Partition the sample space into disjoint events $A_1, A_2, \ldots, A_k \subset \Omega$
- For every scenario i,

The Bayes theorem

- Partition the sample space into disjoint events $A_1, A_2, \ldots, A_k \subset \Omega$
- For every scenario i,
  - we have initial beliefs, $P(A_i)$

The Bayes theorem

- Partition the sample space into disjoint events $A_1, A_2, \ldots, A_k \subset \Omega$

- For every scenario i,
  - we have initial beliefs, $P(A_i)$
  - we observe the event under the scenario, $P(B|A_i)$

The Bayes theorem

- Partition the sample space into disjoint events $A_1, A_2, \ldots, A_k \subset \Omega$

- For every scenario i,
  - we have initial beliefs, $P(A_i)$
  - we observe the event under the scenario, $P(B|A_i)$

- We wish to revise the belief, given the observation B

$$
\begin{aligned}
P(A_i|B) &= \frac{P(A_i \cap B)}{P(B)} \\
&= \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)}
\end{aligned}
$$

Bayesian approach

"Bayesian" traces back to the 18th century and Rev. Thomas Bayes, who, along with Pierre-Simon Laplace, developed what we now call "Bayes' Theorem".

Bayesian approach

"Bayesian" traces back to the 18th century and Rev. Thomas Bayes, who, along with Pierre-Simon Laplace, developed what we now call "Bayes' Theorem".

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)} \propto P(A_i)P(B|A_i)$$

Bayesian approach

"Bayesian" traces back to the 18th century and Rev. Thomas Bayes, who, along with Pierre-Simon Laplace, developed what we now call "Bayes' Theorem".

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)} \propto P(A_i)P(B|A_i)$$

where

$P(A_i\|B)$	posterior
$P(B\|A_i)$	likelihood
$P(A_i)$	prior

Bayesian approach

"Bayesian" traces back to the 18th century and Rev. Thomas Bayes, who, along with Pierre-Simon Laplace, developed what we now call "Bayes' Theorem".

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)} \propto P(A_i)P(B|A_i)$$

where

$P(A_i\|B)$	posterior
$P(B\|A_i)$	likelihood
$P(A_i)$	prior

Gives intuition to update prior beliefs about A_i, observing data B

Bayes's rule — a use case

You plan to visit a beach on a Saturday afternoon; decide based on the possibility of rain[2].

[2] https://www.youtube.com/watch?v=cqTwHnNbc8g

Bayes's rule — a use case

You plan to visit a beach on a Saturday afternoon; decide based on the possibility of rain[2].

- Chance of rain any day 15%
- On rainy afternoons, cloudy in the morning is 80% of the time.
- Probability of being cloudy is 25%

[2] https://www.youtube.com/watch?v=cqTwHnNbc8g

Bayes's rule — a use case

You plan to visit a beach on a Saturday afternoon; decide based on the possibility of rain[2].

- Chance of rain any day 15%
- On rainy afternoons, cloudy in the morning is 80% of the time.
- Probability of being cloudy is 25%

$$
\begin{aligned}
P(A = \text{Rain}) &= .15 && \text{prior} \\
P(B = \text{Cloudy}) &= .25 && \text{marginal probability} \\
P(B = \text{Cloudy} \mid A = \text{Rain}) &= .80 && \text{likelihood}
\end{aligned}
$$

$$
P(A = \text{Rain} \mid B = \text{Cloudy}) = \frac{.15 \times .80}{.25}
$$

[2] https://www.youtube.com/watch?v=cqTwHnNbc8g

Independence of events

Independence of events

- Intuitive definition $P(A \mid B) = P(A)$—occurrence of B provides no extra information about A

Independence of events

- Intuitive definition $P(A \mid B) = P(A)$—occurrence of B provides no extra information about A
- We then have

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

a more apt definition for independence is below

$$P(A \cap B) = P(B)P(A),$$

Independence of events

- Intuitive definition $P(A \mid B) = P(A)$—occurrence of B provides no extra information about A
- We then have

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

a more apt definition for independence is below

$$P(A \cap B) = P(B)P(A),$$

  - Symmetric with A, B

Independence of events

- Intuitive definition $P(A \mid B) = P(A)$—occurrence of B provides no extra information about A
- We then have

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

a more apt definition for independence is below

$$P(A \cap B) = P(B)P(A),$$

  - Symmetric with A, B
  - Implies $P(B \mid A) = P(B)$

Independence of events

- Intuitive definition $P(A \mid B) = P(A)$—occurrence of B provides no extra information about A
- We then have

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

a more apt definition for independence is below

$$P(A \cap B) = P(B)P(A),$$

- Symmetric with A, B
- Implies $P(B \mid A) = P(B)$
- Applies even when $P(A) = 0$ or $P(B) = 0$

More on independence of events

Definition

$$P(A \cap B) = P(B)P(A)$$

More on independence of events

Definition

$$P(A \cap B) = P(B)P(A)$$

- We use $A \perp\!\!\!\perp B$ to denote A and B are independent
- Disjoint and independence are different concepts—example

More on independence of events

Definition

$$P(A \cap B) = P(B)P(A)$$

- We use $A \perp\!\!\!\perp B$ to denote A and B are independent
- Disjoint and independence are different concepts—example

Proposition

If $A \perp\!\!\!\perp B$, then so $A \perp\!\!\!\perp B^{\complement}$

More on independence of events

Definition

$$P(A \cap B) = P(B)P(A)$$

- We use $A \perp\!\!\!\perp B$ to denote A and B are independent
- Disjoint and independence are different concepts—example

Proposition

If $A \perp\!\!\!\perp B$, then so $A \perp\!\!\!\perp B^{\complement}$

Proof. Hint: $A = (A \cap B) \cup (A \cap B^{\complement})$

Independence of a collection of events

Definition

- Intuition: information on some events doesn't change probabilities related to the rest of the events.

Independence of a collection of events

Definition

- Intuition: information on some events doesn't change probabilities related to the rest of the events.
- $A_1, A_2, \ldots,$ are independent. It implies, for example,

$$P(A_2 \cap A_4) = P(A_2 \cap A_4 \,|\, A_1 \cup (A_3 \cap A_5{}^{\complement}))$$

Independence of a collection of events

Definition

- Intuition: information on some events doesn't change probabilities related to the rest of the events.
- $A_1, A_2, \ldots,$ are independent. It implies, for example,

$$P(A_2 \cap A_4) = P(A_2 \cap A_4 \mid A_1 \cup (A_3 \cap A_5{}^{\complement}))$$

- Formally, we write

$$P(A_1 \cap A_2 \cap \ldots \cap A_m) = P(A_1)P(A_2)\ldots P(A_m),$$

for any distinct choice of indices.

Independence of a collection of events

Definition

- $A_1, A_2, \ldots,$ are independent. We then have

$$P(A_1 \cap A_2 \cap \ldots \cap A_m) = P(A_1)P(A_2)\ldots P(A_m),$$

for any distinct choice of indices

Independence of a collection of events

Definition

- $A_1, A_2, \ldots,$ are independent. We then have

$$P(A_1 \cap A_2 \cap \ldots \cap A_m) = P(A_1)P(A_2)\ldots P(A_m),$$

for any distinct choice of indices

Independence of a collection of events

Definition

- $A_1, A_2, \ldots,$ are independent. We then have

$$P(A_1 \cap A_2 \cap \ldots \cap A_m) = P(A_1)P(A_2)\ldots P(A_m),$$

for any distinct choice of indices

Case $n = 3$: Three events A, B, C, are said to be independent, if we have

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

also, have pairwise independence.

$$P(A \cap B) = P(A)P(B)$$
$$P(B \cap C) = P(B)P(C)$$
$$P(A \cap C) = P(A)P(C)$$

Independence of a collection of events

Definition

- $A_1, A_2, \ldots,$ are independent. We then have

$$P(A_1 \cap A_2 \cap \ldots \cap A_m) = P(A_1)P(A_2)\ldots P(A_m),$$

for any distinct choice of indices

Independence of a collection of events

Definition

- $A_1, A_2, \ldots,$ are independent. We then have

$$P(A_1 \cap A_2 \cap \ldots \cap A_m) = P(A_1)P(A_2)\ldots P(A_m),$$

for any distinct choice of indices

- If events A, B, C are independent, then A will be independent of any event formed from B, C.

Independence of a collection of events

Definition

- $A_1, A_2, \ldots,$ are independent. We then have

$$P(A_1 \cap A_2 \cap \ldots \cap A_m) = P(A_1)P(A_2)\ldots P(A_m),$$

  for any distinct choice of indices

- If events A, B, C are independent, then A will be independent of any event formed from B, C.
- For instance, one can prove that A and $(B \cup C)$ are independent

Conditional independence

Definition

Given event C, the conditional independence is defined as the independence under the law $P(. \,|\, C)$.

$$P(A \cap B \,|\, C) = P(B \,|\, C) P(A \,|\, C)$$

Conditional independence

Definition

Given event C, the conditional independence is defined as the independence under the law $P(. \mid C)$.

$$P(A \cap B \mid C) = P(B \mid C) P(A \mid C)$$

Consider the following scenario

- We know that $A \perp\!\!\!\perp B$

Conditional independence

Definition

Given event C, the conditional independence is defined as the independence under the law $P(.\,|\,C)$.

$$P(A \cap B \,|\, C) = P(B \,|\, C)P(A \,|\, C)$$

Consider the following scenario

- We know that $A \perp\!\!\!\perp B$
- If we were told that C **occurred**, can we say that $A \perp\!\!\!\perp B$?

Conditional independence

Definition

Given event C, the conditional independence is defined as the independence under the law $P(. \,|\, C)$.

$$P(A \cap B \,|\, C) = P(B \,|\, C) P(A \,|\, C)$$

Consider the following scenario

- We know that $A \perp\!\!\!\perp B$
- If we were told that C **occurred**, can we say that $A \perp\!\!\!\perp B$?
- Ans: it depends

Conditioning may affect independence — example

- We assume two *biased coins* A, B, with $P(\text{Head} \mid A) = .9$, $P(\text{Head} \mid B) = .1$

Conditioning may affect independence — example

- We assume two *biased coins* A, B, with $P(\text{Head} \mid A) = .9$, $P(\text{Head} \mid B) = .1$
- Given a coin, we toss independently

Conditioning may affect independence — example

- We assume two *biased coins* A, B, with $P(\text{Head} \mid A) = .9$, $P(\text{Head} \mid B) = .1$
- Given a coin, we toss independently

- Are the coin tosses independent, given some additional information (e.g., based on an observation such as the number of heads in the previous n tosses)?